SANTA CLARA UNIVERSITY
Information Systems & Analytics Department

Data Analytics with Python

Syllabus, Fall 2024

Instructor: Alan Tan, Email: ttan@scu.edu
Class hours: TTh 5:45 PM – 7:20 PM
Class Room: 209 Lucas Hall
Office hours: Tuesday 2:30 to 3:30 PM or by appointment

Description

Data science involves the application of scientific methodologies to extract, understand, and make predictions based on data sets from a broad range of sources. Data analytics requires knowledge and skills from three areas: (i) programming, (ii) math/statistics, and (iii) domain specific expertise. The objective of this course is to teach the programming skills relevant to data science. Students will learn to use a complete set of open source tools for data science in Python, including the Jupyter Notebook, NumPy, Pandas, Seaborn, Scikit-learn and many others. Students will learn skills that cover the various phases of exploratory data analysis: importing data, cleaning and transforming data, algorithmic thinking, grouping, aggregation, reshaping, visualization, time series, statistical modeling, and data exploration and communication of results. The course will utilize data from a wide range of sources and will culminate with a final project and presentation.

Learning objectives

Upon successful completion of the course, students will be able to:

- Understand how to use Python and core data science packages,
- Prepare, combine, and wrangle data sources for analysis,
- Develop solutions to exploratory data science problems.

Literature and Learning Material

Teaching material:

- Even though this course name is "Data Analytics with Python" but we will focus more on 'data analytics' instead of 'Python' (as a language). We will only spend minimal time in class covering the basics of Python; instead, students will learn it by completing **one** of the following two online tutorials.
- Option1:
    - **Codecademy**'s online Python Tutorial (https://www.codecademy.com/learn/learn-python). This is an excellent, self-guided tutorial on the Python programming language.

The Python 2 tutorial "was" **free,** But now needs the 'pro' membership beyond Chapter 3. There will be X days of free period. It is possible to finish the class before the free period expires. Or you can take Python 3, which needs 'pro' membership.

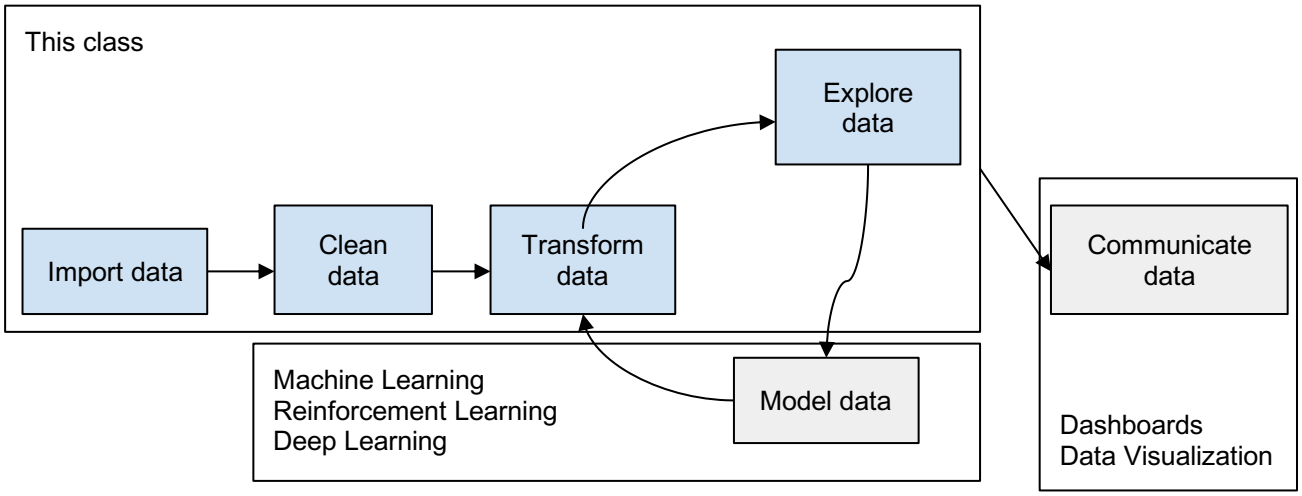
- Option 2:
  - **PY4E, Python for Everybody (**https://www.py4e.com**).** From Chapter 1 to Chapter 10.
  - The tutorial is **free,** but no hands-on interactive interface.

- **<u>You will be required to complete this course by the end of 2nd week. An early start before the quarter is strongly recommended.</u>**
- Textbook(optional) Python for Data Analysis
  - (https://www.amazon.com/Python-Data-Analysis-Wrangling-IPython/dp/1491957662/ref=sr_1_3?keywords=python+data+analysis&qid=1550681598&s=gateway&sr=8-3 , Wes McKinney, O'Reilly (2nd Edition).
- Instructors' slides and lab notes(Jupyter notebook) files.
- Some of the teaching data sets and student's final project data sets can be found in:
  - Kaggle.com: https://www.kaggle.com/
  - Data.gov: https://www.data.gov/
  - Data.world: https://data.world/

Technology

- Camino: SCU's Camino will be used for course materials distribution like homework, project, quizzes, discussions as well as student submissions.
- Laptop: This course will involve extensive programming and computer work, both in and out of class. Students are required to bring a laptop to class each time.
- Anaconda: https://www.anaconda.com . This will be the Integrated Development Environment(IDE) used in this class for the following:
  - Spyder: Python programming
  - Jupyter Notebook: Python open source tools: Pandas, Seaborn, scikit-learn programming

Context



This class focuses on the data preparation and data exploration elements of the data analytics process. As such, this class prepares you for data modeling (e.g., machine learning) and the communication of data (data visualization).

Week 1 to 7	Students will use the content up to week 7 to manually clean up then analyze the data and prepare the data for Machine Learning. In particular, they will be able to perform the task called “feature engineering”, which consists of generating new features that capture information that is not readily available in the given data set.
Week 8 and 9	We will use the data prepared from prior weeks to perform machine learning in classification, prediction, clustering and regression.
Week 10	Wrap up and final project presentation

Course Structure

The course is composed of two parts:

1) Weeks 1-7: Become a “data ninja” (i.e., import, clean, transform data)
2) Weeks 8-10: Become a “data miner” (i.e., explore data)

Week	**Topic**	**Example of tasks that students will learn to**	**Assignment**

		perform	
Wk 1, Getting started	Install the software: The Jupyter notebook environment, Review Python, Numpy	Solve interview-style general coding questions in Python. Like these: https://leetcode.com/problemset/all/	HW1: Based on prior Twitter data, perform a simple analysis.
Wk 2, Pandas	The Series class, The Univariate Time Series, Operating on Data in Pandas, Data Indexing and Selection,	Find what is the expected profit of buying a stock and selling it after x days	HW2: Practice series data manipulation
Wk 3, Pandas	The DataFrame class, Data ingestion (CSV, text, Excel), Hierarchical Indexing, Combining Datasets: Concat and Append	Find the transactions of those clients that made at least 10 purchases	Quiz 1 HW3: Practice Data Frame manipulation
Wk 4, Data cleaning	Data cleaning, Data transformation, Dealing with missing values, Vectorized String Operations,	Adding calculated columns, fill missing values with the average value of the column	Quiz 2
Wk 5, Wk 6, Data wrangling I	Data aggregation, Merging data, Combining Datasets: Merge and Join Midterm exam (1 class)	Compute the total revenue generated by each client, Combine client data with transaction data	HW4: Data manipulation with Groupby, midterm exam
Wk7, Data wrangling II	Advanced Data aggregation: • Aggregation and Grouping • Transposing • Data Formats (Strings, timestamps, etc.) • Transformation wide to long • Pivot table	For each purchases, compute the number of products returned by the same client prior to that purchase. For each client, find the details about his/her most recent product returns.	HW5: Advanced aggregation and shaping Quiz 3
Wk 8, Classification for Data Exploration	Seaborn for data visualization, Simple Line Plots, Simple Scatter Plots Building and interpreting Decision Trees	Plot the total purchases divided by client gender, describe the difference between valuable and non-valuable clients	HW6: Decision trees
Wk9, Introduction into Algorithms for Data Exploration	Classification, Classification for Prediction, Clustering & Regression	Find the groups of homogeneous clients and their characteristics	Quiz 4

Wk10, Wrap up and Team Projects	Project presentations		

Assignments

Match the above-stated learning objectives with the assignments used to grade students.

Learning objective	Assignment	Points	Grade
Learn/refresh Python (as a language)	Online Python Tutorial	10	10%
Practice at home	Home works	20	20%
Key concepts reminders	Quizzes in class	15	15%
Check progress	Midterm exam	20	20%
Get familiar with real-life data	Project	15	15%
Check progress	Final exam	20	20%
Total		100	100%

Grading Scale

Points	Letter
100-95	A
<95-90	A-
<90-87	B+
<87-84	B
<84-80	B-
<80-77	C+

<77-74	C
<74-70	C-
<70	F

Course Policies

- Late homework penalty:
    - 1 day late: -50%
    - More than 1 day: -100%

Statements on Academic Integrity, Office of Accessible Education (Formerly Disabilities Resources), Accommodations for Pregnancy and Parenting, Discrimination and Sexual Misconduct, and In-Class Recordings.

Academic Integrity

The Academic Integrity pledge is an expression of the University's commitment to fostering an understanding of -- and commitment to -- a culture of integrity at Santa Clara University. The Academic Integrity pledge, which applies to all students, states:

> *I am committed to being a person of integrity. I pledge, as a member of the Santa Clara University community, to abide by and uphold the standards of academic integrity contained in the Student Conduct Code.*

Students are expected to uphold the principles of this pledge for all work in this class. For more information about Santa Clara University's academic integrity pledge and resources about ensuring academic integrity in your work, see www.scu.edu/academic-integrity.

Safety Measures

In order to meet our learning objectives, we will adhere to the highest standards for safety and mutual respect. I expect everyone to adhere to current university mask mandates at all times; to make their best attempt to make themselves heard when asking questions or contributing to discussions; and refrain from eating or drinking in class. It is expected that everyone will follow university guidelines about health and public safety measures outlined in this email.

Office of Accessible Education (formerly Disabilities Resources)

If you have a documented disability for which accommodations may be required in this class, please contact the Office of Accessible Education (Benson 1, http://www.scu.edu/oae, 408-554-4109) as soon as possible to discuss your needs and register for accommodations with the University. If you have already arranged accommodations through OAE, please discuss them with me during my office hours within the first two weeks of class.

To ensure fairness and consistency, individual faculty members are required to receive verification from the Office of Accessible Education before providing accommodations. OAE will work with students and faculty to arrange proctored exams for students whose accommodations include double time for exams and/or assistive technology. Students with approved accommodations of time-and-a-half should talk with me as soon as possible. The Office of Accessible Education must be contacted in advance (at least two weeks notice recommended) to schedule proctored examinations or to arrange other accommodations.

Accommodations for Pregnant and Parenting Students

In alignment with Title IX of the Education Amendments of 1972, and with the California Education Code, Section 66281.7, Santa Clara University provides reasonable accommodations to students who are pregnant, have recently experienced childbirth, and/or have medical needs related to childbirth. Pregnant and parenting students can often arrange accommodations by working directly with their instructors, supervisors, or departments. In addition, the Office of Accessible Education will provide reasonable accommodations for pregnancy-related impairments which impact a major life activity.

Discrimination and Sexual Misconduct (Title IX)

Santa Clara University upholds a zero-tolerance policy for discrimination, harassment and sexual misconduct. If you (or someone you know) have experienced discrimination or harassment, including sexual assault, domestic/dating violence, or stalking, I encourage you to tell someone promptly. For more information, please consult the University's Gender-Based Discrimination and Sexual Misconduct Policy or contact the University's EEO and Title IX Coordinator, Belinda Guthrie, at 408-554-3043, bguthrie@scu.edu. Reports may be submitted online through the Office of Student Life https://www.scu.edu/osl/report/ or anonymously through EthicsPoint https://www.scu.edu/hr/quick-links/ethicspoint/

In-Class Recording

The Student Conduct Code (p. 13) prohibits students from "making a video recording, audio recording, or streaming audio/video of private, non-public conversations and/or meetings, inclusive of the classroom setting, without the knowledge and consent of all recorded parties," except in cases of approved disability accommodations. The Student Conduct Code also prohibits the "falsification or misuse, including non-authentic, altered, or fraudulent misuse, of University records, permits, documents, communication equipment, or identification cards and government-issued documents." Dissemination or sharing of any classroom recording without the permission of the instructor would be considered "misuse" and, therefore, prohibited. Violations of these policies may result in disciplinary action by the University. At the instructor's discretion, violations may also have an adverse effect on the student's grade

Wellness and Mental Health Resources

Jesuit education is grounded in concern for the whole person—mind, body, and spirit— and SCU has many resources and programs to support you. Resources that assist with mental

wellness and mindfulness can be found through the Cowell Center and Campus Ministry, to name but a few.

University students may experience stressors or setbacks from time to time that can impact both their academic experience and their personal well-being. These may include academic pressure or challenges associated with relationships, mental health, alcohol or other drugs, identities, finances, etc.

If you are experiencing difficulties, seeking help is a courageous thing to do for yourself and those who care about you. If you are concerned with your progress in this class, please contact me so that we can find solutions together. Drahmann Center can also offer support with issues regarding your academic progress more broadly. For personal concerns, SCU offers many resources, some of which are listed on the Cowell Center website.